Genet. Res., Camb. (1979), **34**, *pp.* 173–188
With 6 text-figures
Printed in Great Britain

A mechanism for RNA–RNA splicing and a model for the control of gene expression

By VINCENT MURRAY and ROBIN HOLLIDAY
National Institute for Medical Research, Mill Hill, London NW7 1AA

(*Received* 12 *July* 1979)

SUMMARY

A mechanism for RNA–RNA splicing is proposed. A species of RNA ('splicer' RNA) hybridizes to precursor mRNA across the splice point. This hybridization can be with intron or exon sequences or both. The double-stranded RNA structure precisely indicates to the splicing enzymes the exact location for exon ligation.

A model for the control of gene expression is presented. The regulation of synthesis of different splicer RNAs will also control which precursor mRNA molecules are spliced. The removal of intervening sequences from a precursor mRNA molecule could be both a signal for that molecule to be transported to the cytoplasm and a means of allowing gene expression.

1. INTRODUCTION

It has recently been demonstrated in eukaryotic organisms that the sequence of nucleotides in several structural genes does not have the colinear arrangement expected from polypeptide and RNA sequence analysis. Sequences are present in mature mRNA (exons) which are not continuous in DNA, since they are interrupted by intervening sequences (introns). These introns are removed at some point prior to mRNA formation, leaving the exons in the same order in mRNA as they are in DNA. This discovery has provoked considerable speculation about the possible function of these intervening sequences (Gilbert, 1978; Tonegawa *et al.* 1978; Travers, 1978; Darnell, 1978; Miller, Konkel & Leder, 1978; Doolittle, 1978; Sagano *et al.* 1979; Crick, 1979).

Initially, intervening sequences were discovered by hybridization of mRNA (or a cDNA copy) to DNA, and the 'looped-out' region of the intervening sequence was visualized in the electron microscope. More recently the intron–exon junction DNA sequences have been determined after comparison of the known mRNA sequence with that of the cloned gene. There are a number of eukaryotic genes which are known to possess intervening sequences including globin (Konkel, Tilghman & Leder, 1978; Tiemeier *et al.* 1978), immunoglobulin (Bernard, Hozumi & Tonegawa, 1978; Tonegawa *et al.* 1978); ovalbumin (Catterall *et al.* 1978; Breathneck *et al.* 1978); rDNA (Glover & Hogness, 1977; Wellauer & Dawid, 1977). The number of intervening sequences present in a gene can vary from one (rDNA) to seven (ovalbumin). The eukaryotic viruses adenovirus 2 (Berget,

0016-6723/79/2828-7590 $01.00

Moore & Sharp, 1977; Chow *et al.* 1977; Akusjarvi & Pettersson, 1979), polyoma (Legon *et al.* 1979), SV40 (Aloni *et al.* 1977; Ghosh *et al.* 1978) and retroviruses (Cordell *et al.* 1978) also splice their genes. Up to the present time the genes which have been examined for the presence of intervening sequences are clearly a non-random sample since they are either genes coding for abundant products or repetitive genes. There is evidence that intervening sequences are transcribed in the mouse β-globin gene (Tilghman *et al.* 1978), an early adenovirus 2 mRNA (Blanchard *et al.* 1978) and yeast tRNA (Knapp *et al.* 1978). In the latter two cases a cell-free extract is capable of excising the intervening sequences. In this paper we will assume that splicing occurs at the RNA level.

Splicing must be an extremely precise process since the polypeptide sequence is accurately conserved in individual proteins. This accuracy might be achieved by three basic types of mechanism. First, a protein could specifically recognize the RNA sequence to be spliced; second, *intramolecular* RNA hybridization could provide a structural signal for the exact point of splicing; or third, *intermolecular* RNA hybridization could provide the signal. Unless there are a large number of different proteins involved in splicing, the first mechanism implies that there are common RNA sequences at exon–intron junctions. However, only a short 'consensus' sequence is present at intron–exon junctions. It is apparent that the actual 'consensus' sequence itself cannot be recognized because these sequences occur elsewhere in mRNA without splicing taking place (Catterall *et al.* 1978). Tonegawa *et al.* (1978) and Ghosh *et al.* (1978) proposed that intramolecular RNA hybridization occurred at the splice point (but base pairing was non-perfect) and this structure was recognized by the splicing enzymes. The drawbacks of this second mechanism are that the exons are not in the correct stereochemical position for ligation to occur and also similar structures could not be formed from sequences of other known intron–exon junctions (Catterall *et al.* 1978; Breathneck *et al.* 1978). A specific three-dimensional structure could occur at the splice point, which is recognized by splicing enzymes. However, for the three-dimensional structure to be stable it would have to be based on secondary structure. Thus, this mechanism also suffers from the drawback that secondary structure is not universally found at intron–exon junctions. For these and other reasons we prefer the third possibility and propose an intermolecular RNA hybridization mechanism in which the necessary specificity and stereochemical orientation for splicing is provided simply by the hybridization of a separate complementary molecule of RNA. In this manner a common enzyme, or group of enzymes, could accomplish splicing of all transcripts. Moreover, the existence of complementary RNA sequences can provide a basis for the control of gene expression in higher organisms.

In the model we describe we assume that heterogeneous RNA (hnRNA) is the precursor of mRNA. hnRNA in the nucleus of higher eukaryotes has a short half-life of 20–30 min (Davidson & Britten, 1973). mRNA is much shorter ($0{\cdot}5$–10×10^3 nucleotide bases, average 3×10^3) than hnRNA ($0{\cdot}5$–30×10^3 bases, average 13×10^3) (Lewin, 1975*a*; Perry *et al.* 1976), and has a sequence complexity

about 5–10 fold less than hnRNA (Lewin, 1975*b*; Perry *et al.* 1976). Analysis of the hybridization kinetics of DNA from higher organisms has shown that about 75% of the genome is composed of interspersed non-repetitive DNA and repetitive DNA (Davidson & Britten, 1973; Bonner *et al.* 1973). The widespread occurrence of this arrangement in higher eukaryotes suggests that it fulfils an important function. Britten & Davidson's (1969) theory of gene expression proposed that the repetitive sequences contained the control elements and the non-repetitive DNA the structural genes. They suggested that a number of control element sequences were adjacent to the gene and that its expression was controlled by a complementary RNA species, which hybridized to the control sequence and induced transcription. One difficulty of this theory is that RNA is unlikely to bind to double-stranded DNA. The mechanism proposed here is in part based on Britten and Davidson's argument, but it does not suffer from this particular difficulty, since the control element ('splicer RNA') binds to single-stranded RNA.

2. A MECHANISM FOR RNA–RNA SPLICING

The primary transcript of DNA is assumed to be hnRNA, which contains both intron and exon sequences in a colinear copy of DNA. In cells producing the protein in question, a species of RNA (splicer RNA) is present and this hybridizes to hnRNA across the splice point. This double-stranded RNA molecule brings the ends of two exons into the correct stereochemical position for phosphodiester bond formation (see Fig. 1). It indicates, by its intrinsic structure, the exact sequence location for RNA breakage and ligation. If base-pairing is perfect for a specified number of bases, perhaps as few as 20, the double-stranded RNA will be recognized by the splicing enzymes and the intron is removed. Splicer RNA could be bound to an enzyme complex which carried out the ligation of exons and also possibly the further steps in processing.

There are three possible ways that splicer RNA can bind at an intron–exon junction (see Fig. 1). If hybridized to exon sequences (Fig. 1*a*) it would, in evolutionary terms, preserve the exon nucleotide sequence neighbouring the splice point. Also it could possibly provide a proofreading step to check the accuracy of mRNA. If splicer RNA hybridized to intron sequences (see Fig. 1*c*), this would not impose any sequence restrictions on the structural genes. As we will explain, it would also provide a basis for the conservation of the intervening sequence. A further possibility is that complementary splicer RNAs binds to both exon and intron sequences (see Figs. 1*b*, 2).

The structure in Fig. 1(*b*) is directly analogous to the likely intermediate in DNA recombination (Holliday, 1964; Sigal & Alberts, 1972; Sobell, 1974; Potter & Dressler, 1976). Complete hydrogen bonding is maintained between complementary bases and this establishes the exact position where polynucleotide chains exchange pairing partners. In DNA recombination it is believed that either pair of strands of like polarity can be cut and rejoined, but in the model of RNA splicing only the hnRNA strands must be susceptible to enzymic cleavage. The distinction between the hnRNA and splicer RNA strands may depend on post-

synthetic methylation of bases (see below). Alternatively, the structure in Fig. 1 (*b*) may be stabilized in only one of the three possible isomeric forms (see Sobell, 1974). If the hnRNA strands exchanging pairing partners are on the inside (see Fig. 2), it is easy to see how an enzyme could recognize and cut these strands.

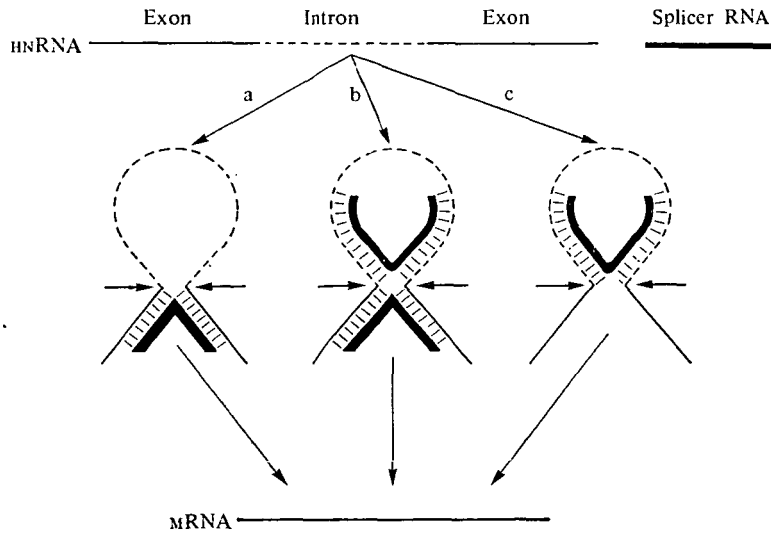

Fig. 1. The hybridization of splicer RNA to hnRNA molecules. In each case the ends of the coding or exon sequences are brought into close juxtaposition. The short arrows indicate the positions where cutting and ligation produces mRNA. Splicer RNA hybridizes to either exon sequences (*a*) or intron sequences (*c*). Two splicer RNA molecules (or one looped back molecule – see Fig. 2) produce a symmetrical structure (*b*) that is the same as the likely intermediate in DNA recombination (see text).

The structures in Fig. 1 (*a*) and (*c*) have only three strands. Fig. 1 (*a*) is like a DNA deletion heteroduplex. There is evidence that such structures can be repaired by removal of the loop and ligation to produce a normal hydrogen-bonded structure (Benz & Berger, 1973; Holliday, 1974). In Fig. 1 (*c*) the hnRNA molecule would be broken and rejoined at the exchange point in the absence of double-stranded exon sequences. Whatever the exact mechanism of the resolution of these structures, the important point is that the paired splicer RNA(s) provides the specificity required for precise joining of exon sequences.

There is, in fact, duplication of a short sequence at each end of every intervening sequence known (Breathneck *et al.* 1978) and in a similar manner to branch migration in DNA recombination, the actual position of the splice point could vary without altering the final mRNA sequence (see Fig. 3).

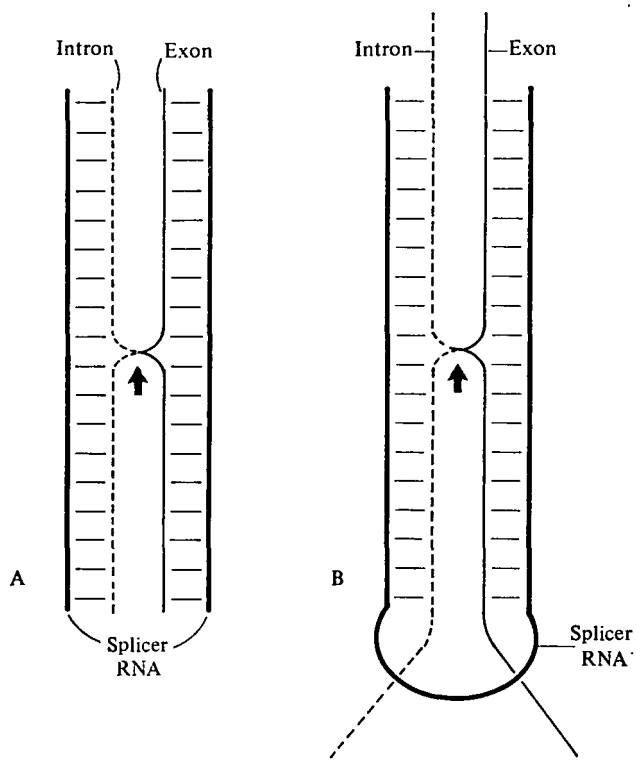

Fig. 2. A possible intermediate in RNA recombination in which the hnRNA strands are on the inside and splicer RNA is on the outside. Cutting and ligation at the point (indicated by an arrow) where hnRNA strands exchange pairing partners generates mRNA. There could be either two splicer RNA molecules (A) or, more simply, one which loops back to form a hairpin structure (B).

3. EVIDENCE FOR RNA–RNA SPLICING MECHANISM

The major prediction of our model is that RNA molecules should be found in the cell nucleus that can hybridize across the splice points of genes expressed in that cell. Direct sequencing of the RNAs involved is the most convincing method for demonstrating this hybridization. In adenovirus 2 infections several small RNAs are produced called VA (virus associated) RNAs which are likely candidates for being splicer RNA. VA RNA_1 is present in large amounts late in infection and has been sequenced (Ohe & Weissman, 1971; Harris & Roeder, 1978). The intron–exon junction sequences for the adenovirus 2 hexon gene (which is expressed late in infection) are known (Akusjarvi & Pettersson, 1979). The third splice-point (from the 5′ end) is shown in Fig. 4, and 19 bases of VA RNA_1 can hybridize across the splice-point in a manner predicted by our splicing mechanism. This hybridization is with both intron and exon sequences (see Figs. 1*b*, 2). This structure, although it contains 4 G-U base pairs, also contains 10 G-C base pairs and therefore should be sufficiently stable for splicing to occur. The probability of this sequence occurring by chance is very remote.

We feel that this sequence hybridization is very strong evidence in favour of our mechanism.

Other evidence is also available for the existence of splicer RNA. Flytzanis *et al.* (1978) isolated ribonucleoprotein complexes (Preobrazhensky & Spirin, 1978) containing hnRNA from rat liver nuclei. These complexes also contained

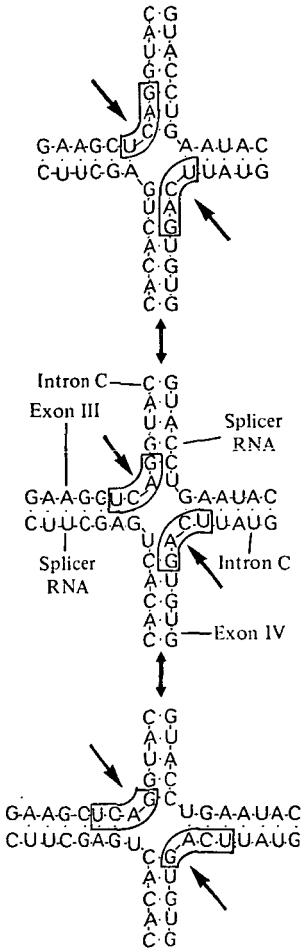

Fig. 3. Three of the five possible splice points connecting exons III and IV in the ovalbumin gene. Duplication of a short sequence (boxed in diagram) at each end of the intron allows branch migration of up to five bases to occur. Splice points are indicated by short arrows.

small nuclear RNA (snRNA) which was base-paired to hnRNA. The length of this hybridization was a minimum of 15–20 nucleotides. snRNA is 150–250 nucleotides in length, highly methylated and with a lifetime greater than 48 h (Sekeris & Niessing, 1978). This stable snRNA is similar in properties to that proposed for splicer RNA. The methylation of snRNA could be the method by which splicer RNA and hnRNA strands are distinguished. The protein complexes, with which snRNA is associated, might contain the splicing enzymes and also possibly

enzymes for the further processing of hnRNA to mRNA. Jelinek & Leinwand (1978) have isolated low-molecular-weight RNA 90–100 nucleotides long from the nuclei and cytoplasm of Chinese Hamster ovary cells. Approximately 1·3 mole of low-molecular-weight RNA was hybridized to 1 mole of nuclear poly(A) terminated RNA but only 0·2 mole was hybridized to 1 mole of cytoplasmic poly(A) terminated RNA.

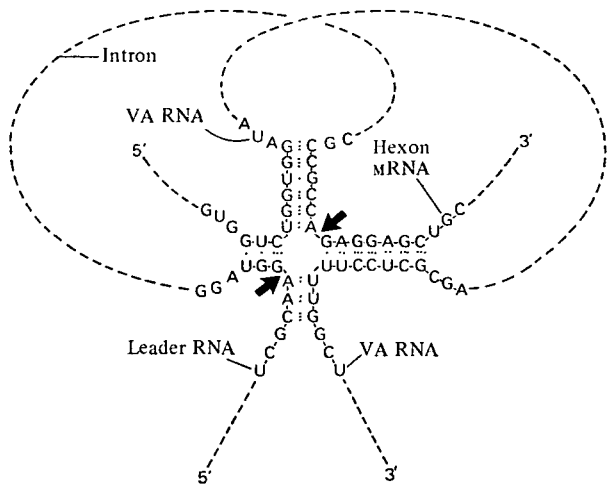

Fig. 4. Hybridization structure formed at the third splice point (from the 5′ end) of adenovirus 2 hexon pre-mRNA with adenovirus 2 VA RNA_1. Two separate segments of VA RNA_1, nucleotides 16–24 and 150–159, hybridize to intron or exon sequences respectively. Arrows indicate splice point.

The sequences of the mouse immunoglobulin light chain genes, λ_1 and λ_{11}, including the small intervening sequence, have recently been determined (Tonegawa *et al.* 1978; Bernard *et al.* 1978). What is strikingly apparent on comparing the two sequences is that the small intervening sequence is as highly conserved as the protein coding sequences. A splicer RNA could base-pair to a possible 29 continuous bases of both introns on either side of the splice point (see Fig. 5) but could only bind to 7 of the exon. This could be taken as evidence that a splicer RNA/intron hybridization structure (see Fig. 1*c*) is being formed at this splice point.

If the sequences of mouse, rabbit and human β-globin mRNA are compared, base substitutions in all three codon positions are rarer in exons neighbouring the two splice points (Kafatos *et al.* 1977; Konkel *et al.* 1978). This sequence conservation (throughout a long period of evolution) implies that constraints exist on this sequence because a special function is assigned to it. Our model suggests that this sequence constraint is due to the necessity of hybridizing splicer RNA to the exon.

The rabbit and mouse β-globin intron–exon junctions have been sequenced for both the large and small intervening sequences (van den Berg *et al.* 1978). The sequences are most highly conserved near the splice points in both intron and

exon sequences. In a similar manner to the immunoglobulin light chain small intron (see Fig. 5), a putative splicer RNA could have the same sequence in both species. This putative splicer RNA could hybridize to 12 bases of the intron or 28 bases of the exon at the small intervening sequence, and a second splicer

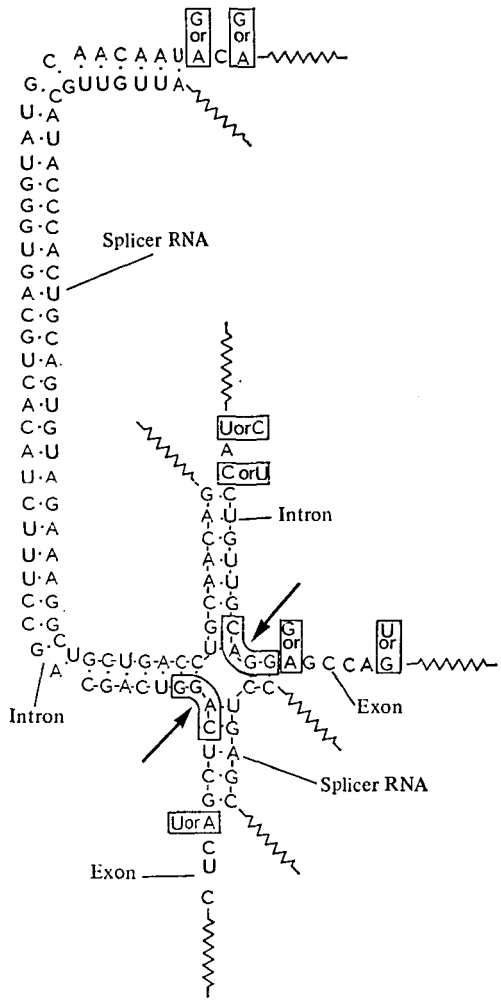

Fig. 5. The splicing of hnRNA transcribed from the mouse immunoglobulin λ_1 and λ_{11} genes. The small intervening sequence is highly conserved (rectangular boxes indicate the only differences in bases). The same splicer RNA(s) could be used to produce the mRNA sequences for both proteins. In the diagram we assume that the hybridization of splicer RNA does not extend past any mismatched base pair.

RNA could hybridize to 10 intron bases or 29 exon bases for the large intervening sequence. Thus splicer RNA hybridizing to exon sequences (see Fig. 1*c*) would seem most likely in this case.

Possible sequence homology between the large intervening sequence of β-major and β-minor mouse globin genes has been examined by restriction enzymes and

heteroduplex mapping analysis (Tiemeier *et al.* 1978). They found that intron sequence homology only existed next to the intron–exon junction.

RNaseP from *Escherichia coli* cleaves tRNA at a specific point near the CCA terminus during maturation of tRNA (Stark *et al.* 1978). It contains RNA, which is essential for its enzymic activity. An interesting possibility is that the enzyme could use its RNA to recognize the exact point of tRNA cleavage (Travers, 1978), in a similar way to the splicing mechanism we have suggested.

4. THE CONTROL OF GENE EXPRESSION

The mechanism for RNA–RNA splicing immediately suggests a simple model for the control of gene expression. 'Splicer RNA' is transcribed from the 'splicer gene'. When a particular splicer RNA is transcribed, hnRNA with a sequence complementary to it will have the appropriate intervening sequence removed. Splicing is an essential step in the maturation of mRNA. If all the intron sequences are removed, the further processing of hnRNA will proceed, including transportation to the cytoplasm, and hence structural gene sequences present on hnRNA will be expressed. hnRNA that is not hybridized to splicer RNA is degraded. Thus, transcription or non-transcription of different species of splicer RNA will decide which genes are expressed.

Transcription of the splicer gene must itself be regulated at the DNA level. We assume this is based on the interaction of specific proteins with promoter or operator sites, possibly depending on the modified or non-modified state of the controlling DNA sequences (see Holliday & Pugh, 1975). We will not review all these possibilities here, but we will argue that transcriptional control of splicer RNA synthesis, combined with control of the processing of hnRNA by splicer RNA provides much greater flexibility in the regulation of gene expression in higher organisms than would either single control mechanism.

The simplest relationship between splicer RNA and gene expression is that every structural gene (A, B, C, etc.) has a single splice point which contains a sequence complementary to the splicer RNAs (a, b, c, etc.). If the transcription of splicer RNAs is specific to particular differentiated cells or tissues, it is apparent that any combination of genes (A, B, C, etc.) can be expressed by controlling the synthesis of the appropriate splicer genes.

However, a more complicated pattern can be obtained if more than one splice point is present in an hnRNA molecule (see Table 1). This would result in a more complex control system using *the same number* of splicer RNA sequences. In fact, the number of possible combinations of genes theoretically controllable by splicer genes increases exponentially. (Number of combinations of genes $= 2^n - 1$, where n = number of splicer sequences.)

In our model the act of removing all the intron sequences from hnRNA is the decisive event which allows an RNA species to be transported to the cytoplasm. There are several molecular mechanisms by which this can be accomplished. For example, every intron could contain an endonuclease-sensitive site. Thus if an intron is not removed, the endonuclease will cleave the RNA and exonucleases

will degrade the rest of the RNA. However, when all the introns are removed, the RNA will be sufficiently long-lived to be transported to the cytoplasm. Alternatively, a covalent modification of the RNA (e.g. methylation) could occur simultaneously with splicing. In this case when all the introns are removed, the RNA will contain the full complement of covalent modifications which are necessary for its transport to the cytoplasm.

Table 1. *Pattern of genes controllable by three splicer RNAs*

Splicer RNAs transcribed	Expressed genes (splicer sequences present) A(a)	B(b)	C(c)	D(a, b)	E(a, c)	F(b, c)	G(a, b, c)
a	✓	—	—	—	—	—	—
b	—	✓	—	—	—	—	—
c	—	—	✓	—	—	—	—
a, b	✓	✓	—	✓	—	—	—
a, c	✓	—	✓	—	✓	—	—
b, c	—	✓	✓	—	—	✓	—
a, b, c	✓	✓	✓	✓	✓	✓	✓

Information in cells is encoded in macromolecules which for the most part cannot move between cells. However, it is possible that splicer RNA could be an important exception to this. Stability in RNA base pairing could readily be achieved by splicer RNA molecules with a molecular weight as low as about 6000 daltons. Molecules of this size might move between cells through intercellular junctions (Subak-Sharpe, Burk & Pitts, 1969; Pitts & Simms, 1977; Kolodny, 1971) or by other means. This would provide a means of cell signalling, where one cell which produced and transmitted splicer RNA would influence gene expression in neighbouring cells.

The theory of Temin (1974) on the origin of RNA tumour viruses suggests that they evolved as aberrations of normal cellular functions. In particular, the transfer of nucleic acids between cells by enveloping RNA in membrane vesicles might be a normal mechanism of information transfer in cell differentiation and development.

5. CONTROL BY PROTEINS OR RNA?

In the proposed mechanism for RNA–RNA splicing, RNA rather than protein provides the necessary specificity for exon ligation. In prokaryotes, gene expression is controlled by proteins binding, or not binding, to specific operator sequences of DNA. A protein that recognizes a specific DNA or RNA sequence must have a precisely constructed binding site. The evolution of such a protein is probably far more difficult than the evolution of an RNA species that can naturally base-pair. In multicellular organisms the number of control elements needed for differentiation and development is much larger than in prokaryotes. Such organisms might therefore evolve the much simpler system of RNA-controlled gene expression. A splicer RNA 20 nucleotides long can contain enough sequence variation for cellular requirements and is capable of extremely complex control of gene expression. In our model the number of possible gene combinations increases

exponentially with the number of splice points and therefore this might explain why more than one intervening sequence is present in genes. (The number of possible different combinations of genes = n^b, where n is the number of splicer RNA sequences and b is the number of introns.)

6. REPETITIVE/NON-REPETITIVE SEQUENCES

The interspersed repetitive/non-repetitive nature of the genome in higher organisms may be explained, at least in part, in this model. Splicer genes, if they are present at many sites on the genome, will behave as repetitive DNA. The part of the intron–exon junction which hybridizes to splicer RNA will also act as repetitive DNA (but will be split in half). Those parts of the exons not involved in splicer RNA hybridization will behave as non-repetitive sequences.

There is evidence for the existence of low-molecular-weight nuclear RNAs analogous to splicer RNA (Davidson & Britten, 1973). These RNAs do not belong to any functionally defined class of RNA. Weinberg & Penman (1968) have designated as RNA species C and D the most common of the long-lived, low-molecular-weight nuclear RNAs. They have been found in all vertebrates tested, are transcribed from repetitive DNA (10^2–10^3 copies per genome), are present in ribonucleoprotein complexes, have a 5′ 'capping' structure and pass into the cytoplasm briefly after their transcription (Eliceiri, 1979).

One prediction of our model is that the intron–exon junction sequences involved in splicer RNA hybridization will also be found elsewhere in the genome. A restriction fragment of the 28 S rDNA intervening sequence has been purified from *D. melanogaster* (Dawid & Botcham, 1977) and *D. virilis* (Barnet & Rae, 1979). In both species this restriction fragment hybridizes to non-rDNA sequences and in the case of *D. melanogaster* this homology comprises 0·2 % of the genome.

Using the Southern hybridization technique, Miller *et al.* (1978) have attempted to find if any of the large intervening sequence of the mouse β-major globin gene is present elsewhere in the genome. They found that sequence homology was present only in the β-minor globin gene. At first sight this data is in disagreement with our model, but in fact the technique is not sensitive enough to detect less than 20–50 base-pair homology, which may well be sufficient for splicer RNA and hnRNA hybridization *in vivo*.

7. REGULATION OF hnRNA TRANSCRIPTION

Estimates of the proportion of the genome being transcribed into hnRNA usually vary between 10 and 20 % (Lewin, 1975*b*). Comparing hnRNA sequences present at different stages of development or in different tissues, one finds that generally less than 80 % homology occurs and in some cases a total lack of homology exists (Davidson & Britten, 1973). It is clear, therefore, that transcriptional control of hnRNA sequences is also taking place. Thus, RNA–RNA splicing is only one type of mechanism for control of gene expression and other forms undoubtedly exist. Transcriptional control could direct gross changes in

the hnRNA population and splicing would exert much finer control in deciding which genes are expressed. Structural genes necessary for 'housekeeping' proteins present in every cell are possibly not controlled by splicing.

8. TRANSPOSON MECHANISM FOR INSERTING INTERVENING SEQUENCES IN STRUCTURAL GENES

The mechanism for RNA splicing requires that intervening sequences are complementary to splicer RNA. This poses the problem of the origin of intervening sequences, which hybridize to splicer RNA, but are positioned at each end of the intron. Fig. 6 shows a possible transposon mechanism by which a controlling sequence of DNA (which already regulates the expression of a structural gene by controlling transcription of the splicer gene) can bring a second structural gene under its control by inserting the splicer gene sequence into the second structural gene. In Fig. 6 the splicer gene is first duplicated, then the copy forms a transposon and inserts itself into the structural gene. It is apparent that sequences complementary to splicer RNA are at each end of the intervening sequence and in the correct position for splicer RNA hybridization to the hnRNA.

Bernard *et al.* (1978) and Breathneck *et al.* (1978) have suggested that the short repeat sequences present at the ends of intron sequences are 'evolutionary relics' of the mechanism by which they were originally inserted. Insertion sequences or transposons in *E. coli* have similar short repeats at their insertion sites.

9. EXPERIMENTAL PREDICTIONS

Our model for RNA splicing makes several experimental predictions. Nuclear ribonucleoprotein complexes ought to contain RNA recombination-like structures when examined in the electron microscope. It also predicts that sequences near the intron–exon junction, which hybridize to splicer RNA, should be conserved during evolution. These sequences should also be found elsewhere in the genome (in splicer genes). In SV40 and adenovirus 2 several of the sequences at the splice points are known and it should be possible to sequence small nuclear RNAs which hybridize to viral hnRNA and our mechanism makes precise predictions about these sequences.

10. CONCLUSIONS

An intermolecular RNA hybridization mechanism for RNA splicing has several advantages. Most importantly, the splice point is accurately determined and the exons are in the correct orientation to be joined by a phosphodiester bond. It also provides a simple reason for the existence of splicing: it functions as a mechanism for the control of gene expression. Evidence is presented from adenovirus 2 that a small RNA of previously unknown function and hexon pre-mRNA can form the proposed RNA recombination structure at the splice point.

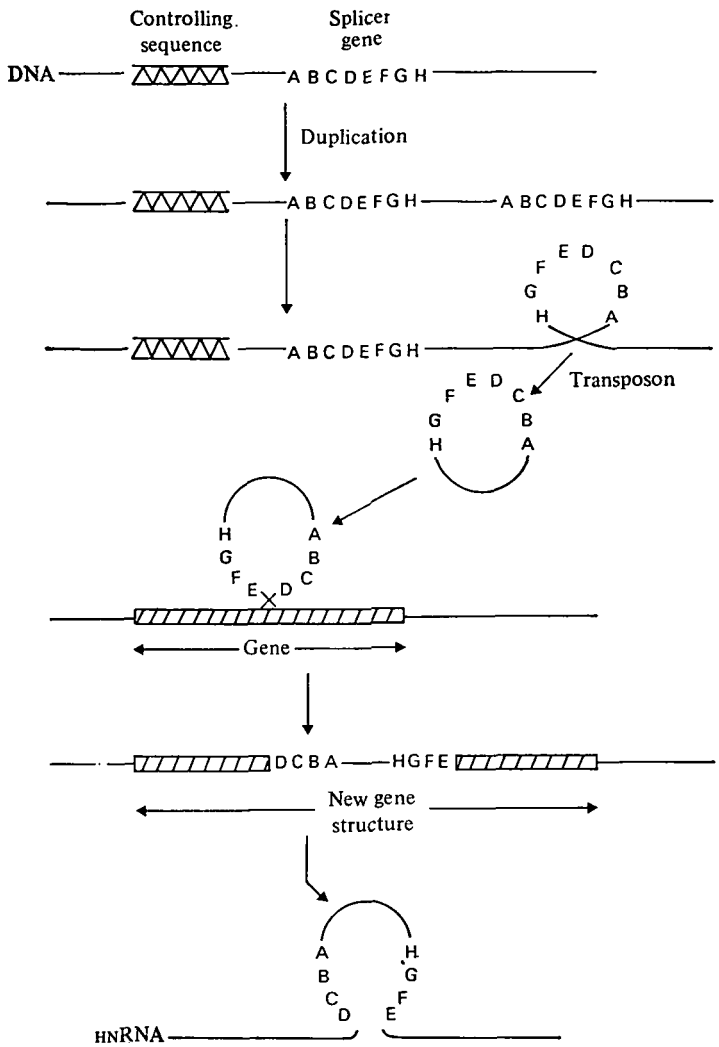

Fig. 6. A transposon mechanism by which intervening sequences that are complementary to splicer RNA can be inserted into genes.

We would like to thank J. Cairns, R. Stellwagen, C. Shaw, J. R. S. Fincham, A. Jeffreys, B. Lewin and H. M. Sobell for helpful comments on this manuscript. V.M. acknowledges support from an M.R.C. studentship.

REFERENCES

Akusjarvi, G. & Pettersson, U. (1979). Sequence analysis of adenovirus DNA: Complete nucleotide sequence of the spliced 5′ non-coding region of adenovirus 2 hexon mRNA. *Cell* **16**, 841–850.

Aloni, Y., Dhar, R., Laub, O., Horowitz, M. & Khoury, G. (1977). Novel mechanism for RNA maturation: The leader sequences of SV40 mRNA are not transcribed adjacent to the coding sequences. *Proceedings of the National Academy of Sciences, U.S.A.* **74**, 3686–3690.

BARNET, T. & RAE, P. M. M. (1979). A 9·6 kb intervening sequence in *D. virilis* r DNA and sequence homology in r DNA interruptions of diverse species of Drosophila and other diptera. *Cell* **16**, 763–775.

BENZ, W. C. & BERGER, H. (1973). Selective allele loss in mixed infections with T4 bacteriophage. *Genetics* **73**, 1–11.

BERGET, S. M., MOORE, C. & SHARP, P. A. (1977). Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences, U.S.A.* **74**, 3171–3175.

BERNARD, O., HOZUMI, N. & TONEGAWA, S. (1978). Sequences of mouse immunoglobulin light chain genes before and after somatic changes. *Cell* **15**, 1133–1144.

BLANCHARD, J.-M., WEBER, J., JELINEK, W. & DARNELL, J. E. (1978). *In vitro* RNA–RNA splicing in adenovirus 2 mRNA formation. *Proceedings of the National Academy of Sciences, U.S.A.* **75**, 5344–5348.

BONNER, J., GARRARD, W., GOTTESFELD, J., HOLMES, D. S., SEVALL, J. S. & WILKES, M. M. (1973). Functional organisation of the mammalian genome. *Cold Spring Harbor Symposium of Quantitative Biology* **38**, 303–310.

BREATHNECK, R., BENOIST, C., O'HARA, K., GANNON, F. & CHAMBON, P. (1978). Ovalbumin gene: Evidence for a leader sequence in mRNA and DNA sequences at the exon–intron boundaries. *Proceedings of the National Academy of Sciences, U.S.A.* **75**, 4853–4857.

BRITTEN, R. J. & DAVIDSON, E. H. (1969). Gene regulation for higher cells: a theory. *Science* **165**, 349–357.

CATTERALL, J. F., O'MALLEY, B. W., ROBERTSON, M. A., STADEN, R., TANAKA, Y. & BROWNLEE, G. G. (1978). Nucleotide sequence homology at 12 intron–exon junctions in the chick ovalbumin gene. *Nature* **275**, 510–515.

CHOW, L. T., GELINAS, R. E., BROKER, T. R. & ROBERTS, R. J. (1977). An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* **12**, 1–8.

CORDELL, B., WEISS, S. R., VARMUS, H. E. & BISHOP, J. M. (1978). At least 104 nucleotides are transposed from the 5′ termini of the Avian Sarcoma Virus genome to the 5′ termini of smaller viral mRNAs. *Cell* **15**, 79–91.

CRICK, F. (1979). Split genes and RNA splicing. *Science* **204**, 264–271.

DARNELL, J. E. (1978). Implications of RNA–RNA splicing in evolution of eukaryotic cells. *Science* **202**, 1257–1260.

DAVIDSON, E. H. & BRITTEN, R. J. (1973). Organisation, transcription and regulation in the animal genome. *Quarterly Review of Biology* **48**, 565–613.

DAWID, I. B. & BOTCHAM, P. (1977). Sequences homologous to ribosomal insertions occur in the Drosophila genome outside the nucleolus organiser. *Proceedings of the National Academy of Sciences, U.S.A.* **74**, 4233–4237.

DOOLITTLE, W. F. (1978). Why genes in pieces? Were they ever together? *Nature* **272**, 581–582.

ELICEIRI, G. L. (1979). Sensitivity of low molecular weight RNA synthesis to U.V. radiation. *Nature* **279**, 80–81.

FLYTZANIS, C., ALONSO, A., LOUIS, C., KRIEG, L. & SEKERIS, C. E. (1978). Association of small nuclear RNA with HN RNA isolated from nuclear ribonucleoprotein complexes carrying HN RNA. *FEBS Letters* **96**, 201–206.

GILBERT, W. (1978). Why genes in pieces? *Nature* **271**, 501.

GHOSH, P. K., REDDY, V. B., SWINSCOE, J., LEBOWITZ, P. & WEISSMAN, S. M. (1978). Heterogeneity and 5′ terminal structures of the late RNAs of SV40. *Journal of Molecular Biology* **126**, 813–846.

GLOVER, D. M. & HOGNESS, D. S. (1977). A novel arrangement of the 18S and 28S sequences in a repeating unit of *Drosophila melanogaster* r DNA. *Cell* **10**, 167–176.

HARRIS, B. & ROEDER, R. G. (1978). Structural relationships of low molecular weight viral RNAs synthesised by RNA Polymerase III in nuclei from adenovirus 2-infected cells. *Journal of Biological Chemistry* **253**, 4120–4127.

HOLLIDAY, R. (1964). A mechanism for gene conversion in fungi. *Genetical Research, Cambridge* **5**, 282–304.

HOLLIDAY, R. (1974). Molecular aspects of genetic exchange and gene conversion. *Genetics* **78**, 273–289.

HOLLIDAY, R. & PUGH, J. E. (1975). DNA modification mechanisms and gene activity during development. *Science* **187**, 226–232.

JELINEK, W. & LEINWAND, L. (1978). Low molecular weight RNAs hydrogen bonded to nuclear and cytoplasmic poly(A) terminated RNA from cultured Chinese Hamster Ovary cells. *Cell* **15**, 205–214.

KAFATOS, F. C., EFSTRADIADIS, A., FORGET, B. A. & WEISSMAN, S. M. (1977). Molecular evolution of human and rabbit β-globin mRNAs. *Proceedings of the National Academy of Sciences, U.S.A.* **74**, 5618–5622.

KNAPP, G., BECKMANN, J. J., JOHNSTON, P. A., FUHRMAN, S. A. & ABELSON, J. (1978). Transcription and processing of intervening sequences in yeast tRNA genes. *Cell* **14**, 221–236.

KOLODNY, G. M. (1971). Evidence for transfer of macromolecular RNA between mammalian cells in culture. *Experimental Cell Research* **65**, 313–324.

KONKEL, D. A., TILGHMAN, S. M. & LEDER, P. (1978). The sequence of chromosomal mouse β-globin major gene: Homologies in capping, slicing and poly(A) sites. *Cell* **15**, 1125–1132.

LEGON, S., FLAVELL, A. J., COWIE, A. & KAMEN, R. (1979). Amplification in the leader sequence of late polyoma virus mRNAs. *Cell* **16**, 373–388.

LEWIN, B. (1975*a*). Units of transcription and translation: the relationship between heterogeneous RNA and mRNA. *Cell* **4**, 11–20.

LEWIN, B. (1975*b*). Units of transcription and translation: sequence components of heterogeneous RNA and mRNA. *Cell* **4**, 77–93.

MILLER, H. I., KONKEL, D. A. & LEDER, P. (1978). An intervening sequence of the mouse β-globin major genes shares extensive homology only with β-globin genes. *Nature* **275**, 772–774.

OHE, K. & WEISSMAN, S. M. (1971). The nucleotide sequence of a low molecular weight RNA from cells infected with adenovirus 2. *Journal of Biological Chemistry* **246**, 6991–7001.

PERRY, R. P., BARD, E., HAMES, B. D., KELLEY, D. E. & SCHIBLER, U. (1976). The relationship between heterogeneous RNA and mRNA. *Progress in Nucleic Acid Research and Molecular Biology* **19**, 275–292.

PITTS, J. D. & SIMMS, J. W. (1977). Permeability of junctions between animal cells. *Experimental Cell Research* **104**, 153–163.

POTTER, H. & DRESSLER, D. (1976). On the mechanism of genetic recombination: electron microscopic observation of recombinant intermediates. *Proceedings of the National Academy of Sciences, U.S.A.* **73**, 3000–3004.

PREOBRAZHENSKY, A. A. & SPIRIN, A. S. (1978). Informosomes and their protein components: The present state of knowledge. *Progress in Nucleic Acid Research and Molecular Biology* **21**, 1–38.

SAKANO, H., ROGERS, J. H., HUPPI, K., BRACK, C., TRAUNECKER, A., MAKI, R., WALL, R. & TONEGAWA, S. (1979). Domains and the hinge region of an immunoglobulin heavy chain are encoded in separate DNA segments. *Nature* **277**, 627–633.

SEKERIS, C. E. & NIESSING, J. (1975). Evidence for the existence of a structural RNA component in the nuclear ribonucleoprotein particles containing heterogeneous RNA. *Biochemical and Biophysical Research Communications* **62**, 642–650.

SIGAL, N. & ALBERTS, B. (1972). Genetic recombination: The nature of a crossed strand exchange between two homologous DNA molecules. *Journal of Molecular Biology* **71**, 789–793.

SOBELL, H. M. (1974). Concerning the stereochemistry of strand equivalence in genetic recombination. In *Mechanisms in Recombination* (ed. R. F. Grell), pp. 433–438. New York: Plenum.

STARK, B. C., KALE, R., BOWMAN, E. J. & ALTMAN, S. (1978). Ribonuclease P: An enzyme with an essential RNA component. *Proceedings of the National Academy of Sciences, U.S.A.* **75**, 3717–3721.

SUBAK-SHARPE, H., BURK, R. R. & PITTS, J. D. (1969). Metabolic cooperation between biochemically marked mammalian cells in tissue culture. *Journal of Cellular Science* **4**, 353–367.

TEMIN, H. M. (1974). On the origin of RNA tumour viruses. *Annual Review of Genetics* **8**, 155–174.

TIEMEIER, D. C., TILGHMAN, S. M., POLSKY, F. I., SEIDMAN, J. G., LEDER, A., EDGELL, M. H. & LEDER, P. (1978). A comparison of two cloned mouse β globin genes and their surrounding and intervening sequences. *Cell* **14**, 237–245.

TILGHMAN, S. M., CURTIS, P. J., TIEMEIER, D. C., LEDER, P. & WEISSMAN, C. (1978). The intervening sequence of a mouse β-globin gene is transcribed within the 15S β-globin mRNA precursor. *Proceedings of the National Academy of Sciences, U.S.A.* **75**, 1309–1313.

TONEGAWA, S., MAXAM, A. M., TIZARD, R., BERNARD, O. & GILBERT, W. (1978). Sequence of a mouse germ-line gene for a variable region of an immuno-globulin light chain. *Proceedings of the National Academy of Sciences, U.S.A.* **75**, 1485–1489.

TRAVERS, A. (1978). RNA processing. *Nature* **275**, 365.

VAN DEN BERG, J., VAN OOYEN, A., MANTEI, N., SCHAMBOCK, A., GROSVELD, G., FLAVELL, R. A. & WEISSMAN, C. (1978). Comparison of cloned rabbit and mouse β-globin genes showing strong evolutionary divergence of two homologous pairs of introns. *Nature* **276**, 37–44.

WEINBERG, R. A. & PENMAN, S. (1968). Small molecular weight monodisperse nuclear RNA. *Journal of Molecular Biology* **38**, 289–304.

WELLAUER, P. K. & DAWID, I. B. (1977). The structural organisation of ribosomal DNA in *Drosophila melanogaster*. *Cell* **10**, 193–212.